C L A I M S

1. A method for the segmentation of an audio stream into semantic or syntactic units wherein the audio stream is provided in a digitized format, comprising the steps of:

determining a fundamental frequency for the digitized audio stream;

detecting changes of the fundamental frequency in the audio stream;

determining candidate boundaries for the semantic or syntactic units depending on the detected changes of the fundamental frequency;

extracting at least one prosodic feature in the neighborhood of the candidate boundaries;

determining boundaries for the semantic or syntactic units depending on the at least one prosodic feature.

2. The method according to claim 1, wherein providing a threshold value for the voicedness of the fundamental

frequency estimates and determining whether the voicedness of fundamental frequency estimates is lower than the threshold value.

3. The method according to claim 2, wherein defining an index function for the fundamental frequency having a value = 0 if the voicedness of the fundamental frequency is lower than the threshold value and having a value = 1 if the voicedness of the fundamental frequency is higher than the threshold value.

4. The method according to claim 3, wherein extracting at least one prosodic feature in an environment of the audio stream where the value of the index function is equal 0.

5. The method according to claim 4, wherein the environment is a time period between 500 and 4000 milliseconds.

6. The method according to claim 1, wherein the at least one prosodic feature is represented by the fundamental frequency.

7. The method according to claim 1, wherein the extracting step involves extracting at least two prosodic features

and combining the at least two prosodic features.

8. The method according to claim 1, further comprising first detecting speech and non-speech segments in the digitized audio stream and performing the steps of claim 1 thereafter only for detected speech segments.

9. The method according to claim 8, wherein the detecting of speech and non-speech segments comprises utilizing the signal energy or signal energy changes, respectively, in the audio stream.

10. The method according to claim 1, further comprising the step of performing a prosodic feature classification based on a predetermined classification tree.

11. An article of manufacture comprising a computer usable medium having computer readable program code means embodied therein for causing segmentation of an audio stream into semantic or syntactic units, wherein the audio stream is provided in a digitized format, the computer readable program code means in the article of manufacture comprising computer readable program code means for causing a computer to effect:

determining a fundamental frequency for the digitized

audio stream;

detecting changes of the fundamental frequency in the audio stream;

determining candidate boundaries for the semantic or syntactic units depending on the detected changes of the fundamental frequency;

extracting at least one prosodic feature in the neighborhood of the candidate boundaries;

determining boundaries for the semantic or syntactic units depending on the at least one prosodic feature.

12. A digital audio processing system for segmentation of a digitized audio stream into semantic or syntactic units comprising:

means for determining a fundamental frequency for the digitized audio stream,

means for detecting changes of the fundamental frequency in the audio stream,

means for determining candidate boundaries for the

semantic or syntactic units depending on the detected changes of the fundamental frequency,

means for extracting at least one prosodic feature in the neighborhood of the candidate boundaries, and

means for determining boundaries for the semantic or syntactic units depending on the at least one prosodic feature.

13. An audio processing system according to claim 12, further comprising means for generating an index function for the voicedness of the fundamental frequency having a value = 0 if the voicedness of the fundamental frequency is lower than a predetermined threshold value and having a value = 1 if the voicedness fundamental frequency is higher than the threshold value.

14. Audio processing system according to claim 12 or 13, further comprising means for detecting speech and non-speech segments in the digitized audio stream, particularly for detecting and analyzing the signal energy or signal energy changes, respectively, in the audio stream.